ISSN: 2689-1204

Research Article

Journal of Applied Material Science & Engineering Research

Applying Linear and Nonlinear Regression Analysis Models to Study the Possible Relationships between Cancer's Risk and CVD/Stroke Risk over a 12-Year Period for a Type 2 Diabetes Patient Based on GH-Method: Math-Physical Medicine (No. 551)

Gerald C Hsu

EclaireMD Foundation, USA

***Corresponding author**
Gerald C Hsu, EclaireMD Foundation, USA

Submitted: 01 Dec 2021; **Accepted:** 08 Dec 2021; **Published:** 20 Dec 2021

Note: *Readers who want to get a quick overview can read the abstract, results and graphs sections.*

Citation: *: Gerald C Hsu (2021) Applying Linear and Nonlinear Regression Analysis Models to Study the Possible Relationships between Cancer's Risk and CVD/Stroke Risk over a 12-Year Period for a Type 2 Diabetes Patient Based on GH-Method: Math-Physical Medicine (No. 551). J App Mat Sci & Engg Res, 5(4), 01-07.*

Abstract

In the author's previous medical research reports, he mainly applied physics theories, engineering models, mathematical equations, computer big data analytics and artificial intelligence (AI) techniques, as well as some statistical approaches to explore and interpret various biophysical phenomena. However, the majority of medical research papers he has read thus far are primarily based on statistics. As a result, in this article, he selects some basic statistical tools, such as correlation, variance, p-values, and ***regression analyses****, to study the predicted Cancer risk probability as the output (dependent variable) by using his MI-based CVD/Stroke risk as the input (independent variable).*

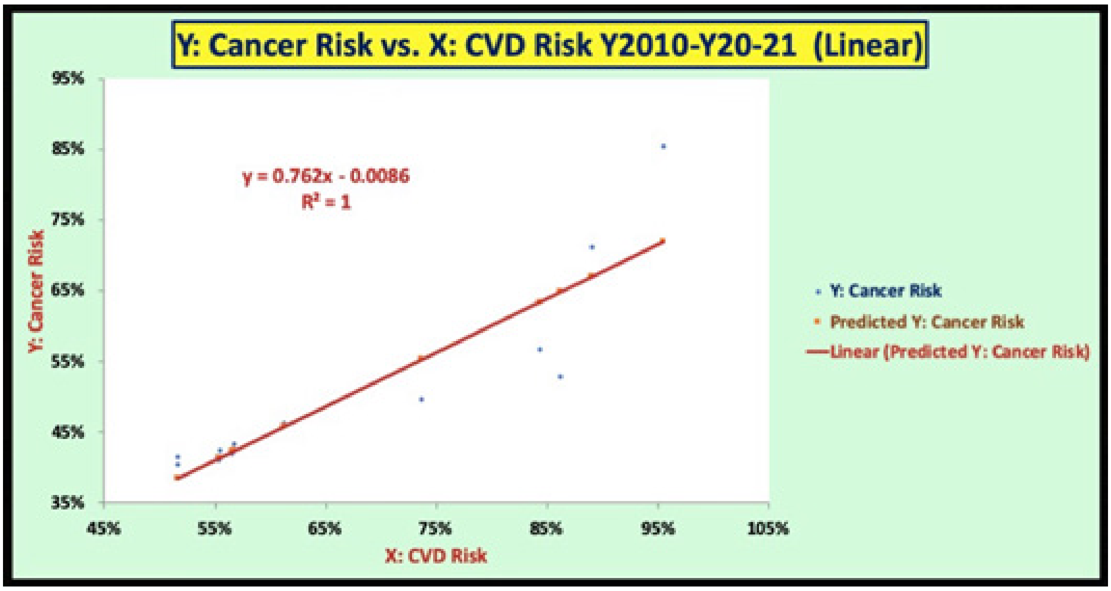
Since 1/1/2012, the author has been collecting various data related to his health (~3 million data) which includes 4 categories of medical conditions, obesity, diabetes, hypertension, and hyperlipidemia (m1 through m4), along with 6 categories of lifestyle details, including exercise, water intake, sleep, stress, food, and daily life routines (m6 through m10). Due to his limited knowledge in earlier years, the datasets from 2010 to 2012 are incomplete; therefore, the data used in this study for the initial period of 2010-2012 are his best-guessed data.

Previously, he researched and published a few articles regarding the risks of having CVD/Stroke and Cancers based on his enhanced metabolism index (MI) model. In this particular paper, adopting a regression analysis model, he is able to compare the previously calculated Cancer risks based on the enhanced MI model versus the newly regression predicted Cancer risks using his MI-based CVD risks as input.

In this study, he will not repeat the detailed introduction of the regression analysis in the Method section because it is available in many statistics textbook. It should be noted that in regression analysis, the correlation coefficient R should be > 0.5 or 50% to indicate a strong inter-connectivity and the p-value should be < 0.05 to be considered as statistically significant.

In summary, there are two specific conclusions worth mentioning:

1. *The time-domain analysis results have revealed that the correlation between his Cancer risk and CVD risk are very high (89%). In addition, his enhanced MI-based model Cancer Risk and Regression Predicted Cancer Risk are highly correlated as well.*
2. *The space-domain linear regression analysis has shown that there is a variance of 1.0 existing between his Cancer risk and CVD risk. When using the nonlinear polynomial model, his variance still reaches 99.6%. This finding has further proven his Cancer risk and CVD risk are highly correlated together. Furthermore, he has conducted two more linear regression analyses of Cancer risk vs. Lifestyle and CVD risk vs. Lifestyle which reveal two extremely high linear variances at 1.0. These findings offer additional indication that Lifestyle, the common root-cause, contributes highly on both CVD risk and Cancer risk.*



Introduction

In the author's previous medical research reports, he mainly applied physics theories, engineering models, mathematical equations, computer big data analytics and artificial intelligence (AI) techniques, as well as some statistical approaches to explore and interpret various biophysical phenomena. However, the majority of medical research papers he has read thus far are primarily based on statistics. As a result, in this article, he selects some basic statistical tools, such as correlation, variance, p-values, and **regression analyses**, to study the predicted Cancer risk probability as the output (dependent variable) by using his MI-based CVD/Stroke risk as the input (independent variable).

Since 1/1/2012, the author has been collecting various data related to his health (~3 million data) which includes 4 categories of medical conditions, obesity, diabetes, hypertension, and hyperlipidemia (m1 through m4), along with 6 categories of lifestyle details, including exercise, water intake, sleep, stress, food, and daily life routines (m6 through m10). Due to his limited knowledge in earlier years, the datasets from 2010 to 2012 are incomplete; therefore, the data used in this study for the initial period of 2010-2012 are his best-guessed data.

Previously, he researched and published a few articles regarding the risks of having CVD/Stroke and Cancers based on his enhanced metabolism index (MI) model. *In this particular paper, adopting a regression analysis model, he is able to compare the previously calculated Cancer risks based on the enhanced MI model versus the newly regression predicted Cancer risks using his MI-based CVD risks as input.*

In this study, he will not repeat the detailed introduction of the regression analysis in the Method section because it is available in many statistics textbook. It should be noted that in regression analysis, the correlation coefficient R should be > 0.5 or 50% to indicate a strong inter-connectivity and the p-value should be < 0.05 to be considered as statistically significant.

Methods

MPM Background

To learn more about the author's developed GH-Method: math-physical medicine (MPM) methodology, readers can select the following three papers from his ~500 published medical papers.

The first paper, No. 386 describes his MPM methodology in a general conceptual format. The second paper, No. 387 outlines the history of his personalized diabetes research, various application tools, and the differences between biochemical medicine (BCM) approach versus the MPM approach. The third paper, No. 397 depicts a general flow diagram containing ~10 key MPM research methods and different tools.

In particular, paper No. 453 illustrates his GH-Method: math-physical medicine in great details, "Using Topology concept of mathematics and Finite Element method of engineering to develop a mathematical model of Metabolism in medicine in order to control various chronic diseases and their complications via overall health conditions improvement".

His Case of T2D and Complications

The author has been a severe type 2 diabetes (T2D) patient since 1996 and weighed 220 lbs. (100 kg, BMI 32.5) at that time. By 2010, he still weighed 198 lbs. (BMI 29.2) with an average daily glucose of 250 mg/dL (HbA1C of 10%). During that year, his triglycerides reached to 1161 (diabetic retinopathy or DR) and albumin-creatinine ratio (ACR) at 116 (chronic kidney disease or CKD). He also suffered five cardiac episodes within a decade. In 2010, three independent physicians warned him regarding his needs of kidney dialysis treatment and future high risk of dying from severe diabetic complications. Other than cerebrovascular disease (stroke), he has suffered most known diabetic complications, including both macro-vascular and micro-vascular complications.

In 2010, he decided to launch his self-study on endocrinology, diabetes, and food nutrition in order to save his own life. During

2015 and 2016, he developed four prediction models related to diabetes conditions: weight, postprandial plasma glucose (PPG), fasting plasma glucose (FPG), and A1C. As a result, from using his developed mathematical metabolism index (MI) model in 2014 and the four prediction tools, by end of 2016, his weight was reduced from 220 lbs. (100 kg, BMI 32.5) to 176 lbs. (89 kg, BMI 26.0), waistline from 44 inches (112 cm, nonalcoholic fatty liver disease /NAFLD) to 33 inches (84 cm), average finger glucose reading from 250 mg/dL to 120 mg/dL, and lab-tested A1C from 10% to ~6.5%. One of his major accomplishments is that he no longer takes any diabetes medications since 12/8/2015.

In 2017, he has achieved excellent results on all fronts, especially his glucose control. However, during the pre-COVID period of 2018 and 2019, he traveled to approximately 50+ international cities to attend 65+ medical conferences and made ~120 oral presentations. This hectic schedule inflicted damage to his diabetes control, through dinning out frequently, post-meal exercise disruption, jet lag, and along with the overall metabolism impact due to his irregular life patterns through a busy travel schedule; therefore, his glucose control and overall metabolism state were somewhat affected during this two-year heavy travel period.

During 2020 and 2021 with a strict COVID-19 quarantine lifestyle, not only has he written and published ~400 medical papers in 100+ journals, but he has also reached his best health conditions for the past 26 years. By the beginning of 2021, his weight was further reduced to 165 lbs. (BMI 24.4) along with a 6.1% A1C value (daily average glucose at 105 mg/dL), without having any medication interventions or insulin injections. These good results are due to his non-traveling, low-stress, and regular daily life routines. Due to the knowledge of chronic diseases, practical lifestyle management experiences, and his developed various high-tech tools, they contributed to his excellent health status since 1/19/2020, which is the start date of being self-quarantine.

On 5/5/2018, he applied a continuous glucose monitoring (CGM) sensor device on his upper arm and checks glucose measurements every 5 minutes for a total of ~288 times each day. He has maintained the same measurement pattern to present day. In his research work, he uses the CGM sensor glucose at time-interval of 15 minutes (96 data per day). Incidentally, the difference of average sensor glucoses between 5-minute intervals and 15-minute intervals is only 0.4% (average glucose of 114.81 mg/dL for 5-minutes and average glucose of 114.35 mg/dL for 15-minutes with a correlation of 93% between these two sensor glucose curves) during the period from 2/19/20 to 8/13/21.

Therefore, over the past 11 years, he could study and analyze the collected ~3 million data regarding his health status, medical conditions, and lifestyle details. He applies his knowledge, models, and tools from mathematics, physics, engineering, and computer science to conduct his medical research work. His medical research work is based on the aims of achieving both "high precision" with "quantitative proof" in the medical findings.

The following timetable provides a rough sketch of the emphasis of his medical research during each stage:

- 2000-2013: Self-study diabetes and food nutrition, developing a data collection and analysis software.
- 2014: Develop a mathematical model of metabolism, using engineering modeling and advanced mathematics.
- 2015: Weight & FPG prediction models, using neuroscience.
- 2016: PPG & HbA1C prediction models, utilizing optical physics, AI, and neuroscience.
- 2017: Complications due to macro-vascular research such as cardiovascular disease (CVD), coronary heart disease (CHD) and stroke, using pattern analysis and segmentation analysis.
- 2018: Complications due to micro-vascular research such as CKD, bladder, foot, and eye issues such as DR.
- 2019: CGM big data analysis, using wave theory, energy theory, frequency domain analysis, quantum mechanics, and AI.
- 2020: Cancer, dementia, longevity, geriatrics, DR, hypothyroidism, diabetic foot, diabetic fungal infection, linkage between metabolism and immunity, and learning about certain infectious diseases such as COVID-19.
- 2021: Applications of LEGT and perturbation theory from quantum mechanics on medical research subjects, such as chronic diseases and their complications, cancer, and dementia. Using metabolism and immunity.it's as the base, he expands his research into cancers, dementia, and COVID-19. In addition, he has also developed a few useful analysis methods and tools for his medical research work.

To date, he has collected nearly 3 million data regarding his medical conditions and lifestyle details. In addition, he has written 536 medical papers and published 500+ articles in 100+ various medical journals, including 7 special editions with selected 20-25 papers for each edition. Moreover, he has given ~120 presentations at ~65 international medical conferences. He has continuously dedicated time and effort on medical research work to share his findings and knowledge with patients worldwide.

Risk Probability Model of CVD

In this paper, the author described how to apply his engineering science background, including mathematics, physics, and computer science to conduct medical research on the risk probability of having various complications from chronic diseases, such as stroke, CVD, CHD, CKD, DK, dementia, and cancer. He will specifically explain the risk model of having a CVD or stroke.

After reviewing 6 years of collected data from 10/12/2016 through 10/11/2021, where he focused on his 4 chronic disease's medical conditions, including obesity, diabetes, hypertension, and hyperlipidemia, along with the medical conditions and 6 lifestyle details.

As a part of his medical research, he applied the acquired mechanical and structural engineering knowledge to develop several biomedical scenarios to research the chronic diseases, obesity, diabetes, hypertension, and hyperlipidemia, along with their induced various complications. One of these complications is CVD/stroke.

The engineering analogy of deaths caused by disease and human expected lifespan can be explained simply by using an example of a new machine or a new bridge. If we develop a monitoring system to continuously measure, record, and analyze the external forces, material strength, and material damages for a machine or bridge, as well as the relationship between *force/stress (similar to causes* of a disease such as lifestyle details) and *deformation/strain (similar to symptoms* of a disease such as medical conditions), we can then have a clear idea how severe the damages are and to determine the useful life or expected lifespan of the machine or bridge.

The author self-studied chronic diseases, metabolism, and food nutrition for 4-years from 2010 to 2013. He started his medical research work by building a mathematical metabolism model in 2014. He named his research methodology as the "GH-method: math-physical medicine (MPM approach)". Over the past 11 years of his medical research work, he has learned that the most important factor is knowing how to apply physics principles and engineering modeling techniques to various biomedical problems. This is different from simply inserting your biomedical data into some existing mathematical equations extended from physical theories and engineering models.

The reason for doing this is that the original mathematical equations associated with the original inventors' theories or models usually come with their inherited boundary conditions. These conditions may or may not fit perfectly with the biomedical situations directly; therefore, you must understand the scope and applicability of these physical theories and engineering models first, and then find a suitable way to apply them. In other words, by learning other people's wisdom first and then find a way to apply their wisdom to your own biomedical problem is the most practical way to solve these different biomedical problems.

The author's simple numerical calculation of risk probability is based on his knowledge and applications of physics law/concept and engineering modeling technique, big data analytics, and his developed mathematical metabolism model. It has depicted a viable way to extend lowering the risk probability of having various complications through an effective metabolic condition improvement and lifestyle maintenance program. This practical method has already been applied and proven effectively in controlling his T2D and its various complications without taking medications for the past 6 years. In his disease risk model, it includes basic conditions, such as genetic and life-time unhealthy habits. When dealing with medical conditions, for example, the artery blockage situations are mainly related to glucose and blood lipids, while artery fracture conditions are primarily related to glucose and blood pressure. Furthermore, his 6 lifestyle details, food & diet, water intake, exercise, sleep, stress, and daily life routines also play vital roles in the determination of CVD/Stroke risks.

Risk Probability Model of Cancers

Medicine is one of the most complex subjects which includes many aspects other academic subjects do not have. For example, the ambiguity between root cause (input) and symptom (output) can be switched under certain circumstances. However, this role-switching phenomena are not frequently observed in the engineering world. For example, metabolic disorder induced chronic diseases are the symptoms of poor lifestyles (root-cause). Complications such as cardiovascular disease or CVD is the outcome of chronic diseases, such as diabetes (input disease). Furthermore, both diabetes and CVD share some common root causes, such as obesity. Furthermore, even cancers and CVD are somewhat related. Certain cancer cases are also related to CVD, diabetes, obesity, and poor lifestyle. Some clinical data have shown that cancer patients are at higher risk of dying from heart disease and stroke.

The above descriptions offer blended pictures of causes and symptoms. These findings have further demonstrated these phenomena of "role-switching or role-changing".

From the 2017 annual US study of death cases, three major causes are chronic diseases with various complications (50%), ***cancers (29%),*** infectious diseases (11%), along with non-diseases related death (10%).

Cancer is an exceedingly difficult and complicated disease that can affect any organ within the body, where abnormal cells divide and mutate rapidly, destroying healthy normal cells in the process. The possible cause of cancer can result from a combination of many different reasons. The author has dedicated the past decade on researching endocrinology and metabolism. He considers that both endocrinology and cancer are quite similar from the viewpoint of "digging into a black box of the inner universe within the human body". However, based on his rudimentary understanding of cancer, he also feels that the diseases caused by cancer are probably at least 10 times more complicated than endocrinology. Although he is not an oncology expert, only a patient and research scientist on chronic diseases and metabolism, he has a strong curiosity and motivation in wanting to know more about his own risk probability of having cancers (there are more than 50 different types of cancers). This reason inspires his cancer research work by using the learned knowledge on metabolism and research methodologies to conduct his own assessment on the relationship between the overall cancer risk versus the overall metabolism.

One information to support his findings is that there are 23 cancer factors causing a total of 45.2% of entire cancer cases in China (around 2.3 million cases per year). Most of these 23 influential factors happened to be a part of the components identified in his developed mathematical model of metabolism.

Figure 2 demonstrates ***six different cancer influential factors*** *from a summarized cancer table developed by the author which connect certain cancer causing or influencing factors and organ systems affected by cancer.* ***This is an ongoing project.*** He still has a lot to learn about cancer diseases, for example, which organs in one particular organ system are most likely to be affected by an influential factor. Therefore, this article only serves as the beginning of his long journey in cancer research using his developed GH-Method: math-physical medicine.

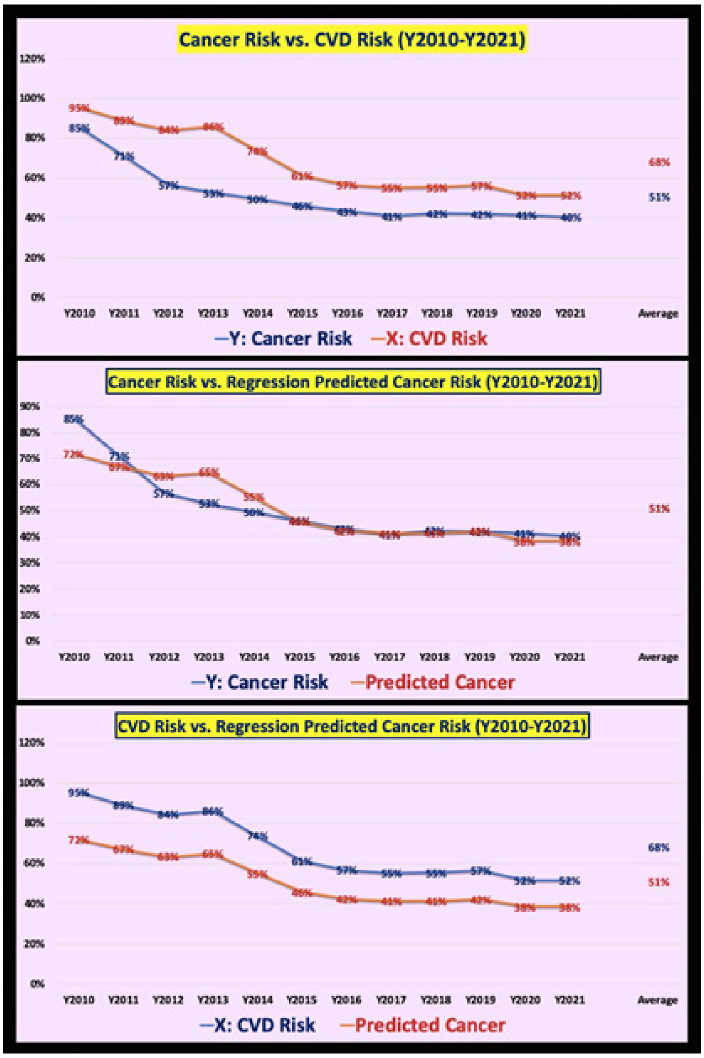


Figure 2: Time-domain analysis of both MI-based Cancer Risk, MI-based CVD Risk, and regression predicted Cancer Risk using CVD Risk as input (2010-2021)

He started his investigation from identifying major causes and the possible organ systems affected by cancer.

First, of course, like many other branches of medical research, he started with the sub-area of genetics, including his age, race, gender, and family genetic background. He has assigned 5% of weight to this sub-area of genetic factors.

Second, he delved into the sub-area of personal unhealthy habits including smoking or chewing tobacco, drinking alcohol, and/or taking illicit drugs that would lead into various types of cancer affecting different organ systems. In addition, he also investigated other components, such as having an inadequate diet, inactive lifestyle, high stress life, poor sleep quality, and personal medical history along with types, amounts, and duration of medication intake that would also lead into different types of cancers. He assigned 20% of weight to this sub-area of personal unhealthy habit factors.

Third, the sub-area of environmental factors includes toxic chemicals, air pollution (e.g., PM 2.5), water pollution, food pollution, poison, hormone therapy, nuclear radiation (e.g., X-ray, CT), UV radiation, infection from parasites and bacteria, or other cancer-causing chemicals, and more. He assigned 15% of weight to this sub-area of environmental factors. As an example, relatively speaking, China can have a higher percentage of cancer cases in this sub-area due to its highly polluted environment, including land, water, and air.

Fourth, the sub-area of viral infection factors includes Helicobacter Pylori, Hepatitis B Virus, Hepatitis C Virus, HIV Virus, Human Papilloma Virus, Epstein-Barr Virus, Paragonimus Sinensis, Human Herpes Virus Type 8, Kaposi's Sarcoma, Hodgkin's Lymphoma, and others. He assigned 10% of weight to this sub-area of viral infection factors.

Fifth, the sub-area of metabolic disorder induced chronic diseases and their various complications include obesity, diabetes, hypertension, hyperlipidemia, CVD, stroke, CKD, bladder infection, hyperthyroidism, bladder infection, foot ulcer, diabetic retinopathy, and more. He assigned 15% of weight to this sub-area of chronic diseases factors.

Sixth, t***he most important sub-area of lifestyle details which count for 35% of the total weight, should be the foundation of the root-causes mentioned in the above situations except for the genetic factor.*** This sub-area include six categories, food and diet, exercise, water intake, sleep, stress, and daily routine life pattern. These categories combined with the fifth sub-area of chronic diseases have approximately 500 detailed elements (from finite "element" method of engineering).

The above-described 6 influential factor groups and assigned weighting scales for cancers are quite different from his risk assessment model of other chronic disease induced complications, such as CVD/Stroke, CKD, and DR, which are mainly based on 2 major influential factor groups, medical conditions and lifestyle details. It should be noted that CVD/Stroke, CVD, and DR have their certain specific individual influential biomarkers included with their individual medical conditions.

The author spent 10 years to develop and continuously enhance a sophisticated and customized software program to collect all types of input data. He then processed them dynamically in order to provide a daily guideline for the purpose of improving his overall metabolism. Once his metabolism is in good condition, then his immune system will be strong enough to defend against the infectious diseases, such as COVID-19.

The above paragraphs have described the backbone of his mathematical model to calculate the risk probability of having cancers, a combination of all different cancer cases, which is closely related to his metabolism model.

Results

Figure 1 displays a summarized data table of the linear regression analyses of his Cancer risk versus his CVD/Stroke risk over a 12-year period from 2012 to 2021. There are 12 observations (years) with the significance F value of 0.00009. The p-value is also at 0.00009 which is <0.05; therefore, ***the analysis results are statistically significant.***

	Y: Cancer Risk	X: CVD Risk
Y2010	85%	95%
Y2011	71%	89%
Y2012	57%	84%
Y2013	53%	86%
Y2014	50%	74%
Y2015	46%	61%
Y2016	43%	57%
Y2017	41%	55%
Y2018	42%	55%
Y2019	42%	57%
Y2020	41%	52%
Y2021	40%	52%
Average	51%	68%
Correlation =	89%	
Variance =	80%	

	Y: Cancer Risk	Predicted Cancer
Y2010	85%	72%
Y2011	71%	67%
Y2012	57%	63%
Y2013	53%	65%
Y2014	50%	55%
Y2015	46%	46%
Y2016	43%	42%
Y2017	41%	41%
Y2018	42%	41%
Y2019	42%	42%
Y2020	41%	38%
Y2021	40%	38%
Average	51%	51%
Correlation =	89%	
Variance =	80%	

	X: CVD Risk	Predicted Cancer
Y2010	95%	72%
Y2011	89%	67%
Y2012	84%	63%
Y2013	86%	65%
Y2014	74%	55%
Y2015	61%	46%
Y2016	57%	42%
Y2017	55%	41%
Y2018	55%	41%
Y2019	57%	42%
Y2020	52%	38%
Y2021	52%	38%
Average	68%	51%
Correlation =	100%	
Variance =	100%	

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.893559352
R Square	0.798448315
Adjusted R Square	0.778293146
Standard Error	0.066112338
Observations	12

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	0.173151162	0.17315116	39.61506522	0.00009
Residual	10	0.043708413	0.00437084		
Total	11	0.216859575			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	-0.008550617	0.084607754	-0.1010619	0.921498938	-0.197068441	0.179967207
X: CVD Risk	0.76200824	0.121068031	6.29404999	0.00009	0.492251856	1.031764624

Figure 1: Data table of regression analysis results of Cancer Risk resulting from CVD Risk (201-2021)

The 5 key data are listed as follows:
Correlation (R) = 89%
Variance (R^2) = 80%
MI averaged risk = 51%
Regression Predicted risk = 51%
Prediction accuracy = 100%

Figure 2 illustrates the comparison of three time-domain curves, MI-based CVD risk curve, MI-based Cancer risk, and his regression predicted Cancer risk curve using CVD risk as input. All of these three time-domain curves have extremely high correlations.

Figure 3 shows the regression predicted Cancer risk using his MI-based CVD risk as input. This space-domain regression model has identified a perfect 1.0 linear variance and 0.9961 nonlinear (exponential) regression variance. This finding proves his Cancer risk is highly correlated with the CVD risk.

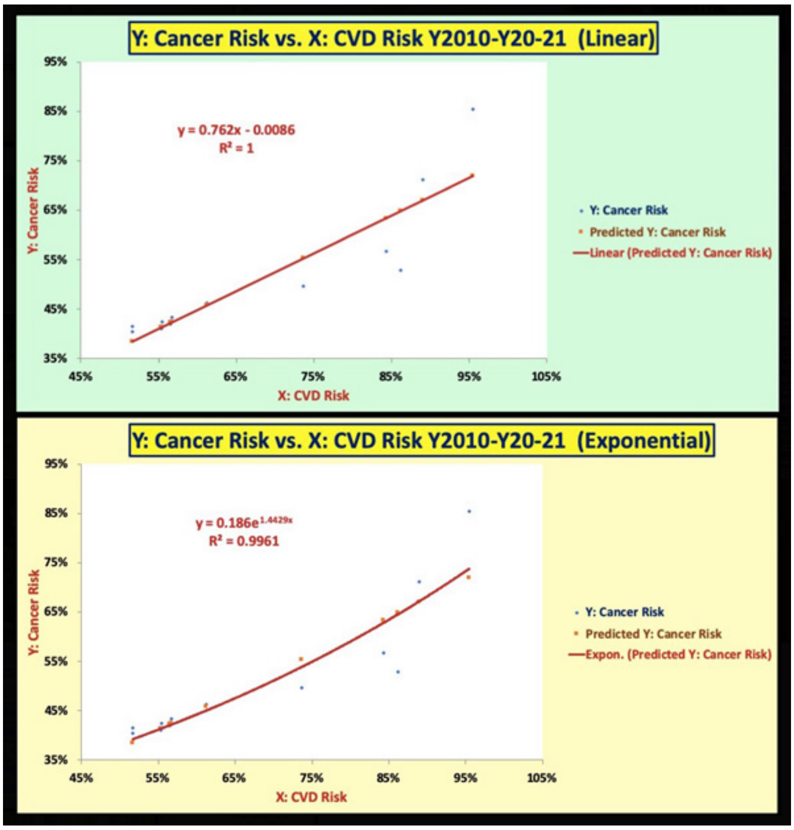


Figure 3: Space-domain regression analysis of Cancer Risk using MI-based CVD Risk as input having 1.0 linear variance (2012-2021)

As a reference, listed below is the linear regression equation for estimating his annual Cancer risk:

Y (predicted Cancer risk)
= 0.762 * X (CVD Risk) - 0.086

And listed below is the nonlinear (exponential) regression equation for estimating his annual Cancer risk:

Y (predicted Cancer risk)

$$= 0186 * e^{1.4429X\ (CVD\ Risk)}$$

Conclusions

In summary, there are two specific conclusions worth mentioning:

1. The time-domain analysis results have revealed that the correlation between his Cancer risk and CVD risk are very high (89%). In addition, his enhanced MI-based model Cancer Risk and Regression Predicted Cancer Risk are highly correlated as well.
2. The space-domain linear regression analysis has shown that there is a variance of 1.0 existing between his Cancer risk and CVD risk. When using the nonlinear polynomial model, his variance still reaches 99.6%. This finding has further proven his Cancer risk and CVD risk are highly correlated together. Furthermore, he has conducted two more linear regression analyses of Cancer risk vs. Lifestyle and CVD risk vs. Lifestyle which reveal two extremely high linear variances at 1.0. These findings offer additional indication that Lifestyle, the common root-cause, contributes highly on both CVD risk and Cancer risk.

References

For editing purposes, majority of the references in this paper, which are self-references, have been removed for this article. Only references from other authors' published sources remain. The bibliography of the author's original self-references can be viewed at www.eclairemd.com.

Readers may use this article as long as the work is properly cited, and their use is educational and not for profit, and the author's original work is not altered.